

# Bioinformatic analysis indicates a relationship between CpG islands & nuclear protein function, as well as a Stop Codon Usage Bias

Paul Berkman<sup>1</sup>  
[pjnhn@hotmail.com](mailto:pjnhn@hotmail.com)

Guang Bin Liu<sup>2</sup>  
[liu@usq.edu.au](mailto:liu@usq.edu.au)

<sup>1&2</sup>: Department of Biological and Physical Sciences, Centre for Systems Biology, Faculty of Science, University of Southern Queensland, Toowoomba 4350, Australia

Key Words: CpG islands, Nuclear protein function, Human chromosome 1, Stop codon

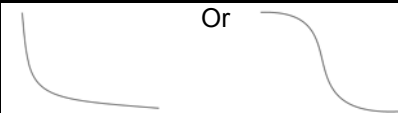
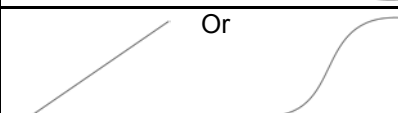
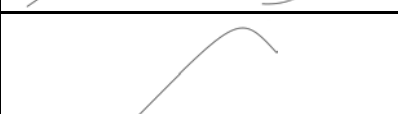
## 1 Introduction

This study utilised pre-existing data, obtained through the National Center for Biotechnology Information (NCBI) and Human Protein Reference Databases (HPRD) [1], to consider distribution of CpG islands (CGIs) and the codon usage bias along in relation to the gene expressions on human chromosome 1.

## 2. Method and results

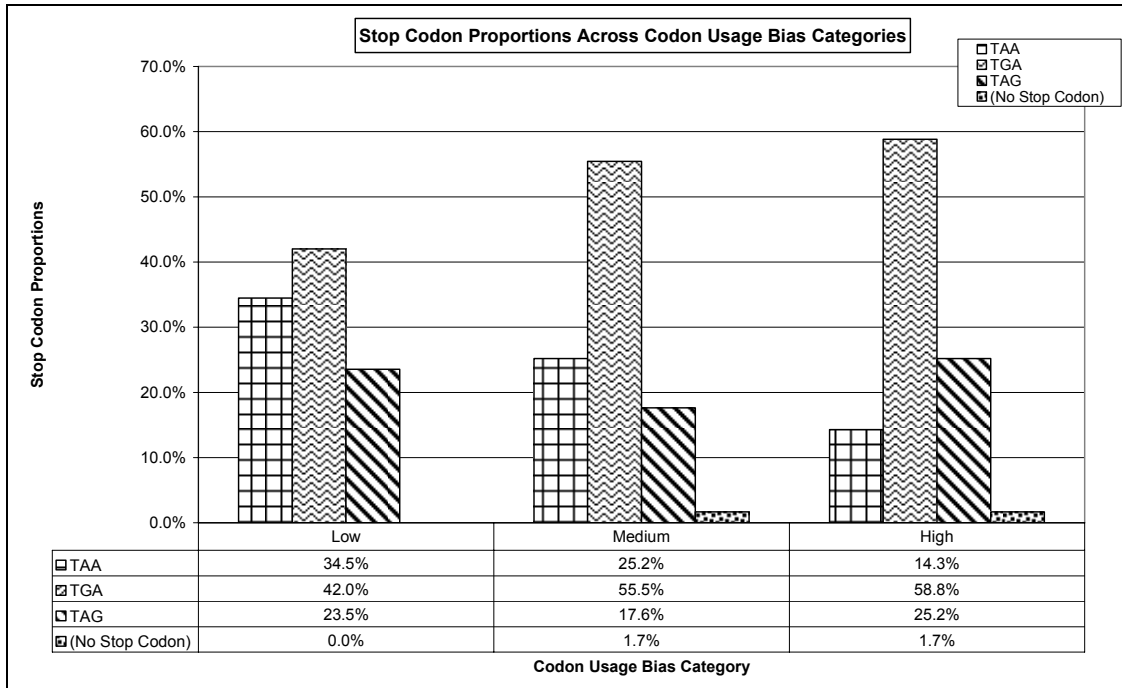
All available genes on human chromosome 1 were obtained from the NCBI Map Viewer [2] and categorised on the basis of the presence of CGIs and density of CpG dinucleotides in the upstream 1kbp, and the extent of codon usage bias. Subsequently, for genes sampled from the resulting categories the gene length and presence of core promoter elements was calculated, with each gene's expressed tissue types, protein size, localisations, domains and motifs obtained from the HPRD. Each of these elements was then compared across the CGI groups using one-way ANOVA and  $\chi^2$  tests, and similarly with the codon usage bias categories.

Table 1 – Trends in Relationship between CGI Density & Expressed Proteins

Trend (% Occurrences vs. CGI Density)	Cellular Localisations	Domains	Motifs
 Or	Plasma membrane, Integral to Membrane	TM, SUSHI	SP
 Or	Nucleus, Nucleolus	ZNFC2, EF	CC
	Cytoplasm	WD40	NLS

A strong relationship was also demonstrated between CGIs and some core promoter elements, particularly the GC Box and TATA Box. The extent of codon usage bias was found to correlate with a bias in stop codon usage, as demonstrated in Figures 1.

Figure 1 – Bar Chart of Stop Codon Proportions across Codon Usage Bias sample groups



### 3 Discussion

The results of this study provide compelling evidence to indicate that the GC Box may have a functional link with CpG islands in the promotion of gene expression. The results also strongly indicate that CGI density in the upstream promoter region of genes is related to the localisation and function of proteins within the cell, in particular that nuclear and cytoplasmic proteins proportionally increase with the presence of a CGI in the upstream promoter region of the gene while membrane protein genes appear to be expressed at inverse proportion to the density of CGIs in the upstream promoter region.

#### References:

1. Peri, S. et al. (2003), Development of human protein reference database as an initial platform for approaching systems biology in humans, *Genome Res*, 13(10): 2363-71.
2. NCBI Map Viewer (<http://www.ncbi.nlm.nih.gov/mapview>)